

Classical Movie Searching

Jadie Zuo

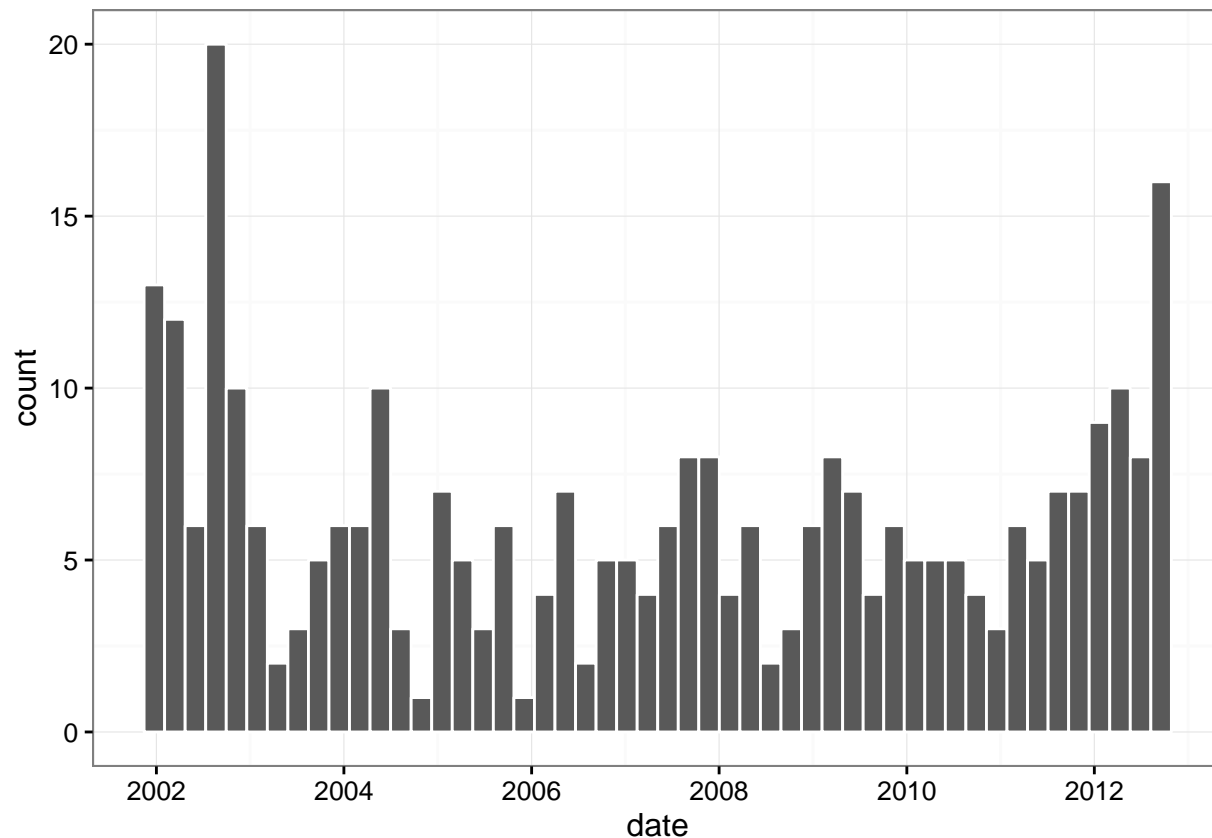
April 4, 2016

Set up

```
library("ggplot2")
library("ggthemes")
library("scales")
library("plyr")
library("grid")
movie <- readRDS("products_100.RDS")
movie1 <- readRDS("users_50.RDS")
movie1$date <- as.Date(as.POSIXct(movie1$review_time,origin="1970-01-01"))
movie$date <- as.Date(as.POSIXct(movie$review_time,origin="1970-01-01"))
movie$year <- format(movie$date,'%Y')
```

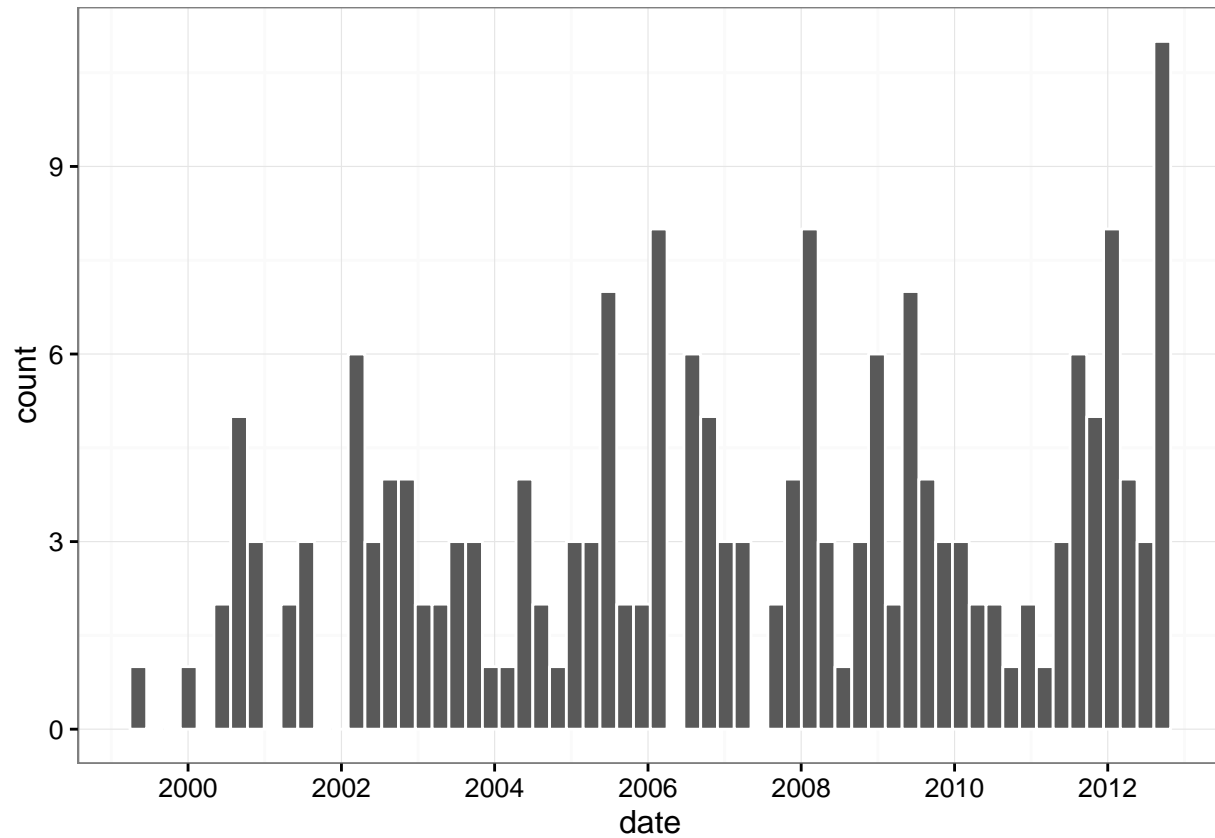
Test 1

```
test1 <- movie[which(movie$product_productid == movie$product_productid[1]),]
p <- ggplot(test1, aes(date, ..count..)) +
  geom_histogram(binwidth = 80,color = "white") + theme_bw()
p
```



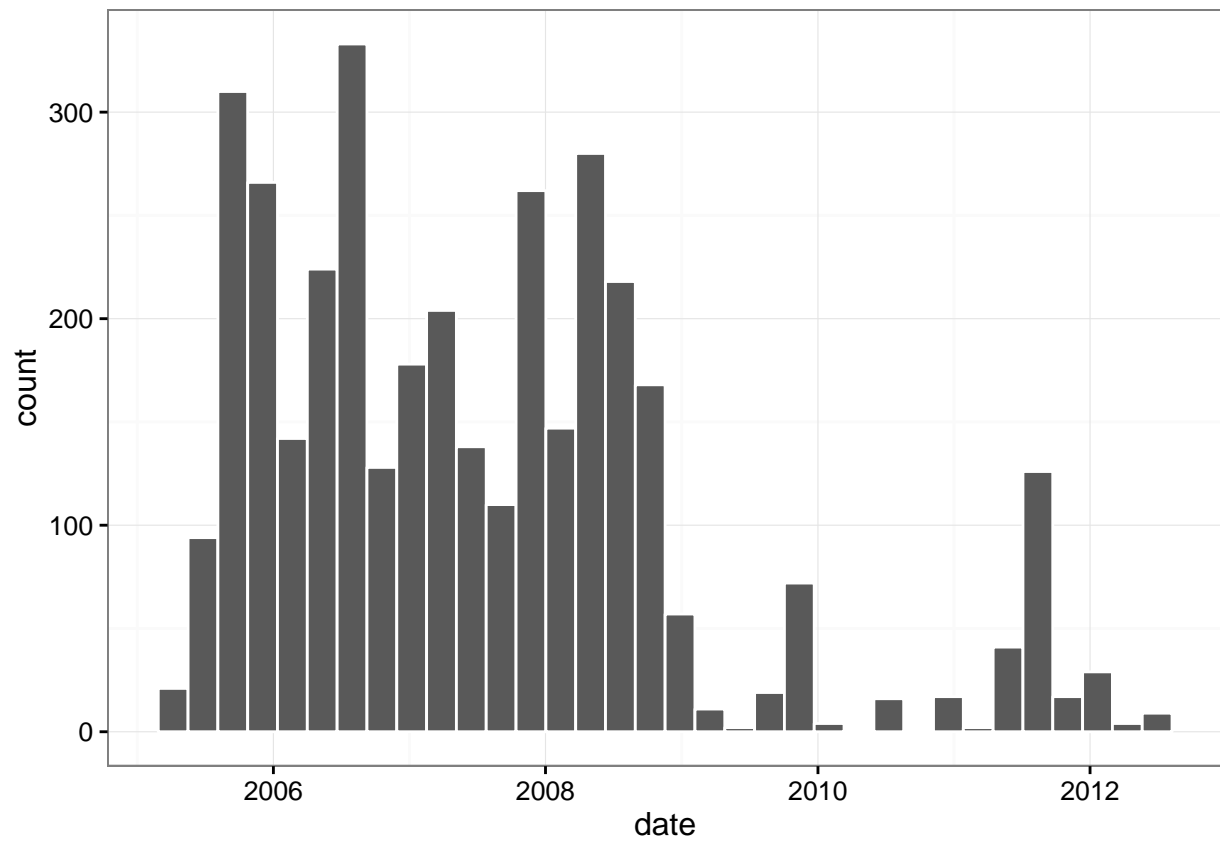
Test 2

```
test2 <- movie[which(movie$product_productid ==  
                      movie$product_productid[dim(test1)[1]+1]),]  
pp <- ggplot(test2, aes(date, ..count..)) +  
  geom_histogram(binwidth = 80,color = "white") + theme_bw()  
pp
```



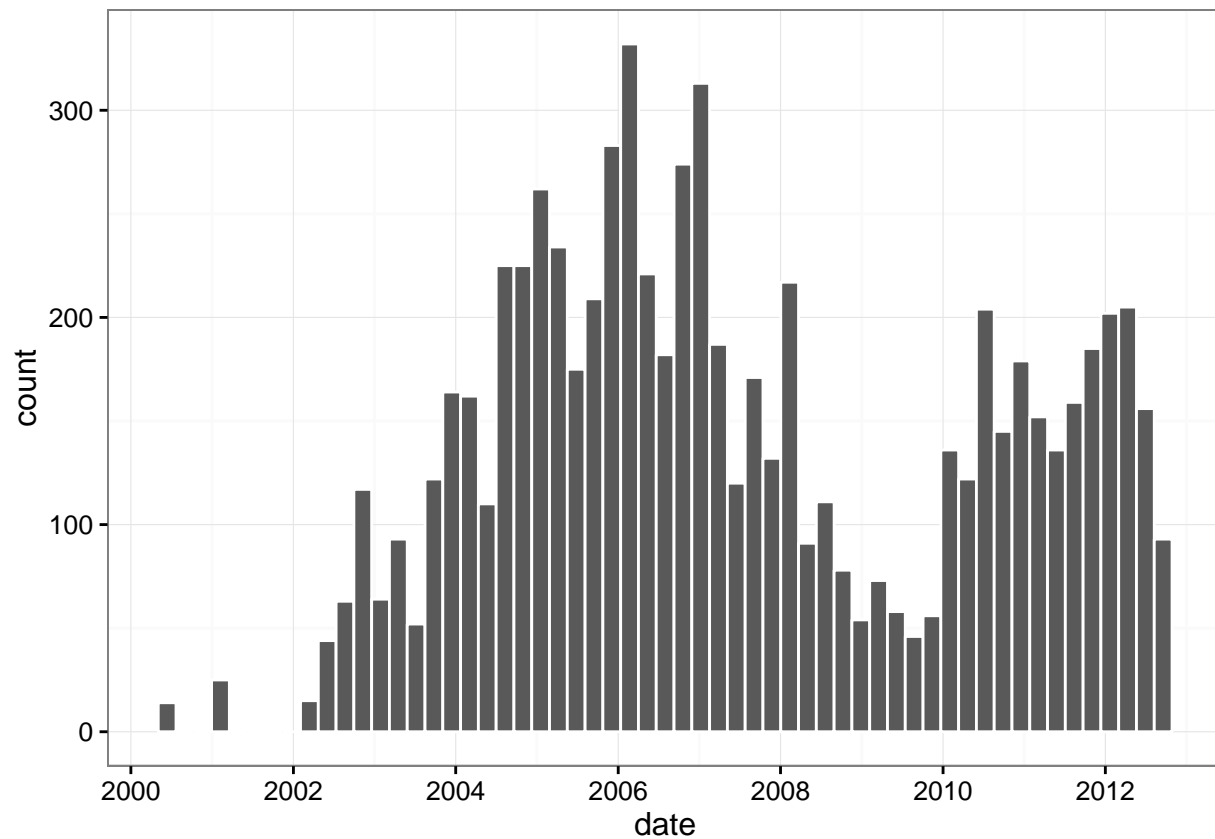
User 1

```
user1 <- movie1[which(movie1$review_userid == movie1$review_userid[1]),]  
u <- ggplot(user1, aes(date, ..count..)) +  
  geom_histogram(binwidth = 80,color = "white") + theme_bw()  
u
```



User 2

```
user2 <- movie1[which(movie1$review_userid == movie1$review_userid[2]),]  
uu <- ggplot(user2, aes(date, ..count..)) +  
  geom_histogram(binwidth = 80, color = "white") + theme_bw()  
uu
```

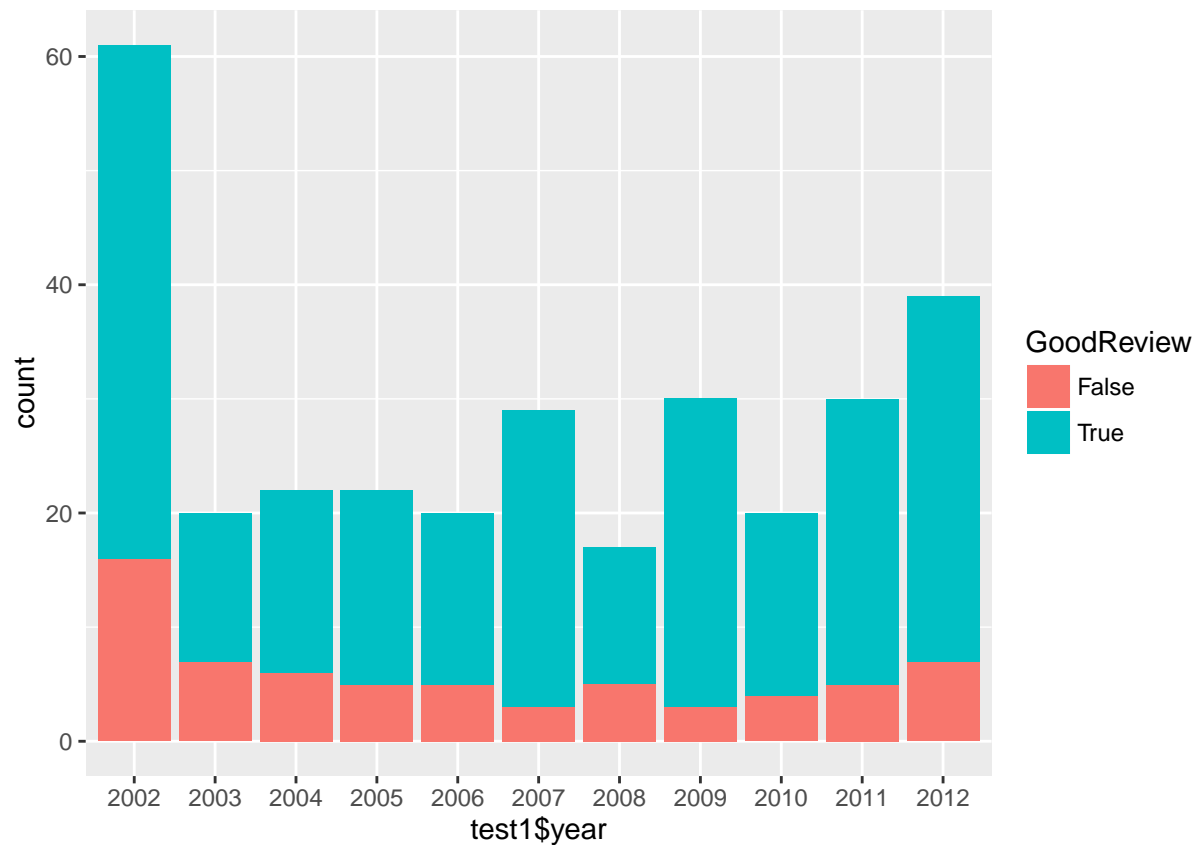


Take a close look at test1

```
test1$year <- format(test1$date, '%Y')
test1$score <- rep("False", dim(test1)[1])
for (i in 1:dim(test1)[1]) {
  if (test1$review_score[i] > 3) {
    test1$score[i] = "True"
  }
}
GoodReview <- factor(test1$score)
```

Plot the review count by years

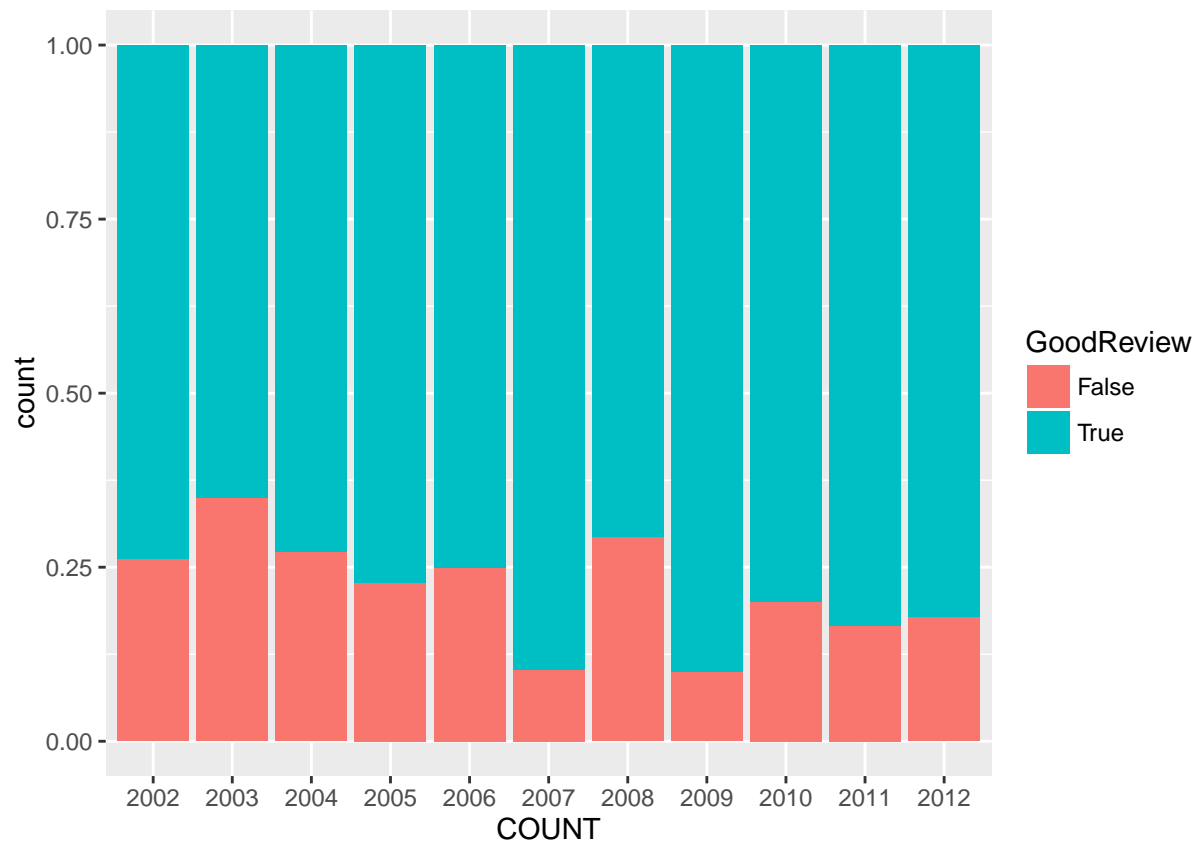
```
ggplot(test1, aes(test1$year, fill=GoodReview)) + geom_bar()
```



Plot

the proportion for reviews with 4 or more stars

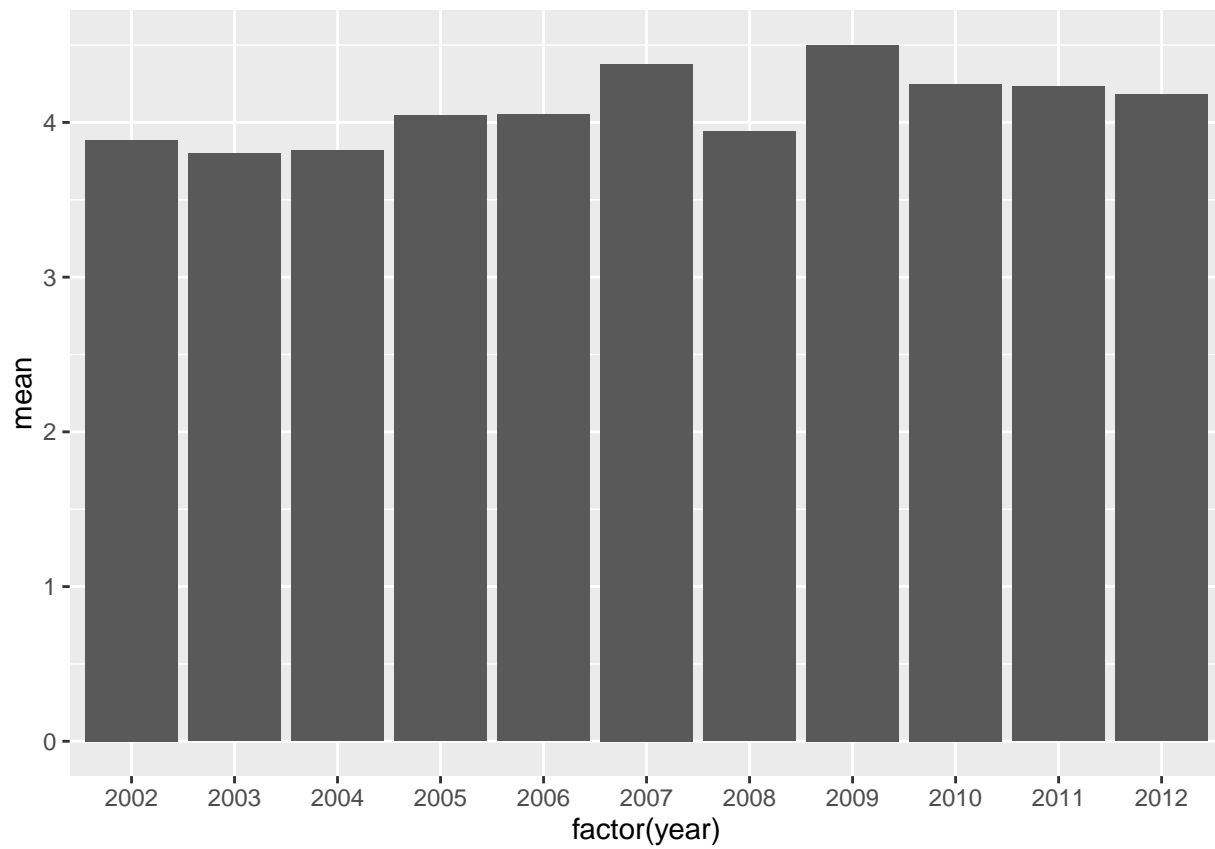
```
ggplot(test1, aes(x=test1$year)) +  
  geom_bar(aes(fill=GoodReview), position="fill") + xlab("COUNT")
```



review scores over years

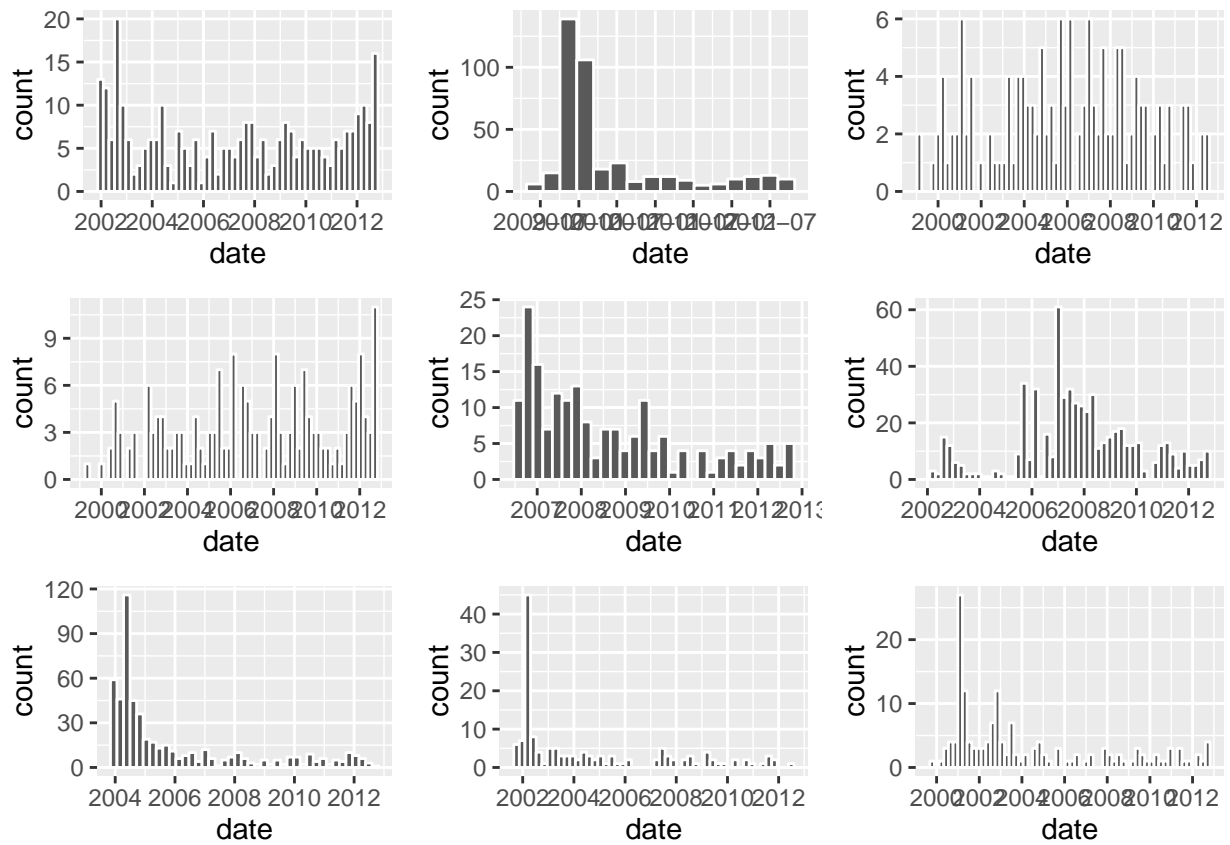
Plot

```
mm <- ddply(test1, .(year), summarize, mean=mean(review_score))
ggplot(mm, aes(x = factor(year), y = mean)) + geom_bar(stat = "identity")
```



Trend plot for the first 9 movies

```
multiplot(p1,p2,p3,p4,p5,p6,p7,p8,p9,cols = 3)
```



```
multiplot(g1,g2,g3,g4,g5,g6,g7,g8,g9,cols = 3)
```